

Intégration relationnelle des exemples lexicographiques dans un réseau lexical

Veronika Lux-Pogodalla
CNRS, ATILF, UMR 7118
Nancy, F-54000, France
Veronika.Lux@atilf.fr

Résumé. Nous présentons un ensemble d'exemples lexicographiques intégré dans le Réseau Lexical du Français et explorons son intérêt potentiel en tant que corpus annoté pour la recherche en désambiguïsation sémantique automatique.

Abstract. This paper presents a set of lexicographic examples which is being developed along the French Lexical Network. The possibility of using this set as an annotated corpus for research on automatic Word Sense Disambiguation is examined.

Mots-clés : Réseau Lexical du Français, exemples lexicographiques, corpus annoté sémantiquement.

Keywords: French Lexical Network, lexicographic examples, semantically annotated corpus.

Objectifs

Le Réseau Lexical du Français (RL-fr) (Lux-Pogodalla & Polguère, 2011) est élaboré par une équipe de lexicographes à l'ATILF-CNRS¹, selon l'approche de la Lexicologie Explicative et Combinatoire (LEC) (Mel'čuk *et al.*, 1995; Mel'čuk, 2006).

Le RL-fr est développé en même temps que différentes ressources linguistiques (ex. ensemble structuré de caractéristiques grammaticales, hiérarchie d'étiquettes sémantiques, base de fonctions lexicales) qui sont utilisées dans les descriptions lexicographiques du RL-fr. L'ensemble, RL-fr et ressources, forment un *système lexical* dont l'architecture est présentée dans (Polguère, 2009) et (Polguère, 2014). Dans ce système, le RL-fr est donc un graphe dont les nœuds, correspondants aux unités lexicales, sont liés aux éléments d'autres structures, les ressources linguistiques.

En cela l'architecture du RL-fr présentent des points communs avec celle de la base FrameNet, dans laquelle trois composants (*Dictionnaire*, *Base de Frames*, *Phrases d'exemples annotés*²) forment un tout complètement intégré et très densément connecté dans lequel les éléments de chaque composant peuvent pointer vers des éléments des deux autres composants³ (Baker *et al.*, 1998). Les données de FrameNet peuvent être consultées sur le site <https://framenet.icsi.berkeley.edu/> en utilisant soit un index de Frames, soit un index des Unités Linguistiques, soit un index des textes annotés. De façon comparable, comme tout le système lexical – le RL-fr proprement dit et les ressources linguistiques – est stocké dans une unique base de données SQL (Gader *et al.*, 2012) et comme les liens sont bidirectionnels, plusieurs perspectives différentes sont possibles dans ce système. Ainsi, les données linguistiques qualifiées de *ressources* lorsque nous nous focalisons sur le RL-fr, peuvent, chacune à leur tour, être vue comme *donnée principale* et placée au premier plan.

Dans cet article, après avoir brièvement décrit le RL-fr dans la section 1, nous nous concentrons, dans la section 2, sur une ressource particulière, la collection d'exemples lexicographiques. Dans la section 3, nous faisons une incursion hors du champ de la lexicographie et opérons un complet changement de perspective, en considérant le réseau lexical comme un *graphe d'annotation* et la collection d'exemples, comme un *corpus annoté*.

L'article se veut une invitation, en particulier à l'adresse des chercheurs en désambiguïsation sémantique automatique, à

1. De juin 2011 à octobre 2014, RL-fr est développé dans le cadre du projet de R&D RELIEF, financé par l'agence de Mobilisation Économique de Lorraine et par le Fonds Européen de Développement Régional

2. Texte source : *Lexicon, Frame Database, Annotated Example Sentences*

3. Texte source : *a highly relational and tightly integrated whole : elements in each may point to elements in the other two*

utiliser ce *corpus* qui, comme tout le RLF, sera bientôt mis à disposition de la communauté. Cette démarche a au moins un précédent couronné de succès, puisque les phrases d'exemples annotés de FrameNet ont été utilisées par différentes équipes pour entraîner des systèmes à base d'apprentissage réalisant de l'annotation automatique de rôles sémantiques (Automatic Semantic Role Labeling)⁴.

1 Le Réseau Lexical du Français RL-fr

Le RL-fr s'appuie sur les recherches menées antérieurement à Montréal à l'Observatoire de linguistique Sens-texte (OLST) par Alain Polguère et Igor Mel'čuk et en particulier sur l'expérience acquise dans le développement :

- de dictionnaires : Dictionnaire explicatif et combinatoire du français contemporain (Mel'čuk et al., 1984 1988 1992 1999), Lexique actif du français (Mel'čuk & Polguère, 2007), et
- de bases de données lexicales : le DiCo (Steinlin *et al.*, 2005) et ses versions en ligne Dicouèbe (<http://olst.ling.umontreal.ca/dicouebe/>) et Dicopop (<http://olst.ling.umontreal.ca/dicopop/>).

Étant un réseau lexical, le RL-fr est membre de la récente famille des ressources lexicales en *-Net*, comme WordNet (Fellbaum, 1998) ou FrameNet (Baker *et al.*, 2003). Mais la nature des nœuds du RL-fr et surtout la variété de ses liens en font une ressource originale, potentiellement une ressource générique à partir de laquelle générer automatiquement différents dictionnaires : c'est pourquoi le RL-fr appartient à la nouvelle génération des « dictionnaires virtuels » (Selva *et al.*, 2003).

Le RL-fr se présente comme un graphe :

- dont les nœuds sont, essentiellement des unités lexicales⁵, soit lexèmes (ex. VIOLON_I) soit locutions (⌈accorder ses violons⌋, ⌈pisser dans un violon⌋), mais aussi des clichés linguistiques (*Arrête ton numéro !*);
- dont les liens sont, essentiellement, des liens de fonctions lexicales (Mel'čuk, 1996), mais aussi des liens entre copolysèmes et des liens d'inclusion formelle (liens entre une locution et ses composants).

Les fonctions lexicales permettent de modéliser des liens paradigmatiques (relations sémantiques entre unités lexicales) ou syntagmatiques (relations de cooccurrence entre unités lexicales). La figure 1 montre quelques liens de fonctions lexicales entrants et sortants pour le nœud FLÛTE_I, par exemple, un lien Real₁⁶ entre FLÛTE_I et JOUER_{IV.1} ainsi qu'entre FLÛTE_I et SOUFFLER_{II}. La figure 2 montre les copolysèmes (c.-à-d. les autres acceptions, dans un vocable polysémique) du nœud FLÛTE_I.

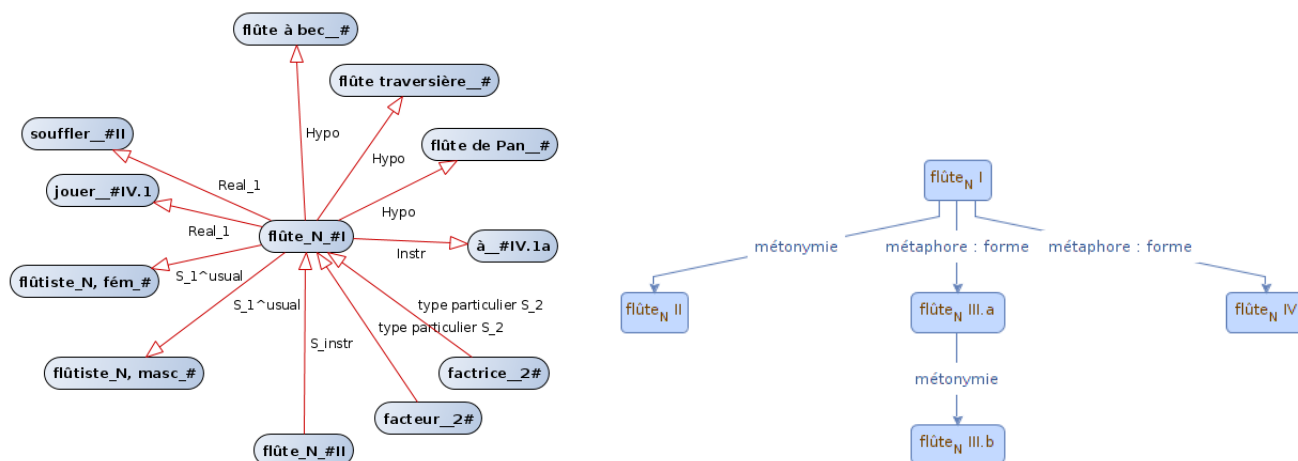


FIGURE 1 – Fonctions Lexicales pour le nœud FLÛTE_I FIGURE 2 – Copolysèmes de FLÛTE_I

Le RL-fr a commencé avec une nomenclature du français fondamental de 3 800 vocables et son expansion est guidée par certaines fonctions lexicales (Polguère & Sikora, 2013). À terme, notre objectif est de couvrir l'ensemble du français courant. Après presque trois ans de développement, le RL-fr compte environ 22 000 lexies (15 000 vocables) et 42 000 liens.

4. Par exemple, le système Shalmaneser développé à Saarbrücken (<http://www.coli.uni-saarland.de/projects/salsa/shal>), le système LTH, développé à Lund (<http://nlp.cs.lth.se/software>) et le système SEMAFOR développé à Carnegie Mellon University (<http://www.ark.cs.cmu.edu/SEMAFOR>)

5. Une unité lexicale correspond à une acception.

6. Real₁ lie une unité lexicale L à un verbe de réalisation dont le premier actant est le premier actant de L et dont le deuxième actant est L.

Le travail lexicographique sur RL-fr est largement manuel, ce qui l'apparente au traditionnel artisanat des dictionnaires papier, mais il s'en différencie pourtant radicalement. Le lexicographe ne rédige pas d'articles : il tisse des liens à l'aide d'un éditeur de graphe (Gader *et al.*, 2012). Ce tissage de liens (Polguère, 2012) entre le RL-fr et les ressources linguistiques de description s'analyse aussi comme une annotation, la ressemblance entre système lexical et graphe d'annotation (Bird & Liberman, 2001) ayant été relevée par (Polguère, 2009). Chaque champ de la description linguistique associée à une unité lexicale du RL-fr correspond à une couche d'annotation de l'unité lexicale avec des éléments d'une structure ressource particulière.

Dans le système lexical, le RL-fr est annoté avec des éléments de toutes les structures ressources. À l'inverse, chaque structure ressource se trouve aussi annotée avec des éléments du RL-fr. Ainsi, dans la base de fonctions lexicales, cette annotation associe des exemples aux fonctions lexicales.

Dans la section suivante, nous présentons une ressource particulière, la collection d'exemples lexicographiques.

2 Collection d'exemples lexicographiques

Selon une pratique aujourd'hui standard dans les projets lexicographiques (Atkins & Rundell, 2008; Kilgarriff *et al.*, 2006), les lexicographes du RL-fr illustrent les unités lexicales qu'ils décrivent avec de nombreux exemples, et, en particulier, avec des exemples *authentiques*⁷. Pour le RL-fr, ces exemples sont cherchés dans un corpus dit *corpus-réservoir*⁸, qui comprend actuellement : des textes de Frantext (<http://www.frantext.fr>) écrits après 1950, des articles de *L'Est Républicain* (presse quotidienne régionale) écrits entre 1999 et 2002 et FrWac (Baroni *et al.*, 2009), une "tranche" du Web de 2009⁹. Les lexicographes du RL-fr illustrent chaque unité lexicale avec au moins neuf exemples du corpus-réservoir¹⁰. Chaque exemple lexicographique est stocké dans la base de données RL-fr avec diverses métadonnées (source, adaptation éventuelle, etc.).

L'ensemble des exemples illustrant des unités lexicales du RL-fr comprend actuellement environ 1 000 000 de mots et 23 000 exemples¹¹. Il est une collection d'extraits soigneusement sélectionnés (Atkins & Rundell, 2008; Selva *et al.*, 2009), dans des textes écrits en français (langue source). Ainsi, les exemples lexicographiques sont courts, généralement composés d'une seule phrase. Un "bon" exemple lexicographique contient une occurrence non ambiguë de l'unité lexicale à illustrer. Autant que possible, l'exemple utilise du vocabulaire "simple" – en particulier, un exemple contenant du vocabulaire rare ou spécialisé est inadéquat pour illustrer une unité lexicale courante. De bons exemples pour l'unité lexicale L contiennent, autant que possible la réalisation des actants sémantiques de L ou bien des cibles de Fonctions Lexicales ; l'ensemble des exemples illustre, idéalement, différentes constructions syntaxiques possibles pour L, même si, dans RL-fr, nous ne cherchons pas à illustrer systématiquement *toute la combinatoire d'une unité lexicale*¹², comme le font les auteurs de FrameNet¹³.

La collection d'exemples du RL-fr est-elle, de par sa composition, très différente des corpus constitués pour la désambiguïsation sémantique automatique ? Elle ne ressemble certes pas au corpus décrit dans (Rakho *et al.*, 2012), qui se compose de phrases, en français langue source ou traduites en français, extraites de compte-rendus de débats parlementaires et contenant des occurrences de 20 verbes polysémiques particuliers du français. Elle ne ressemble par non plus au corpus de néerlandais décrit dans (Vossen *et al.*, 2011), qui est, essentiellement, une juxtaposition de corpus précédemment constitués (dont le corpus SoNaR comptant 500 millions de mots) et comprend des textes entiers plutôt que des phrases isolées. Cependant, ces deux corpus développés pour la désambiguïsation sémantique automatique ne se ressemblent pas entre eux non plus.

7. Soulignons bien que notre démarche n'est pas philologique : nous n'avons pas pour objectif de décrire tous les sens observés en corpus, contrairement à ce qui était le cas dans le *Trésor de la Langue Française* (Dendien & Pierrel, 2003). Aussi notre corpus n'est-il pas à RL-fr exactement ce que Frantext fut au *Trésor de la Langue Française*. Dans notre approche, l'identification des différents sens d'un vocable n'est même pas "dirigée par le corpus", comme elle peut l'être avec un logiciel de traitement de corpus tel que Sketch Engine (Kilgarriff & Kosem, 2012) ; la description de la polysémie repose en premier lieu sur l'intuition du lexicographe. Cependant, si aucun exemple n'est trouvé pour illustrer un sens donné, l'existence de ce sens est évidemment mise en doute.

8. Faute de place, nous n'aborderons pas ici les questions relatives au format des corpus et aux outils de recherche dans ces corpus.

9. FrWac est interrogeable sur : http://nl.ijs.si/noske/wacs.cgi/first_form

10. Ils utilisent aussi des exemples extraits de lectures personnelles (livres, journaux, courriers, Web, etc.) ou entendus à la radio, à la télévision ou dans des conversations, etc. Si nécessaire, les lexicographes adaptent des exemples authentiques, par exemple, en raccourcissant une phrase très longue, en remplaçant un pronom par un groupe nominal, etc.

11. À titre de comparaison : FrameNet compte plus de 170 000 exemples annotés.

12. Texte source : all combinatorial possibilities of the lexical unit

13. Selon : <https://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf>

3 Changement de perspective : le graphe lexical RL-fr comme ressource d'annotation, la collection d'exemples lexicographiques comme corpus annoté

Chaque texte ajouté à la collection d'exemples lexicographiques RL-fr est annoté avec un lien entre un mot-forme du texte et une unité lexicale du RL-fr. Techniquement, cette collection d'exemples est donc un corpus annoté avec un réseau lexical, et notre architecture n'est pas très différente de celle décrite dans (Vossen *et al.*, 2011).

Le RL-fr est formellement un graphe, ce qui l'apparente à d'autres ressources d'annotation utilisées en désambiguïsation sémantique automatique. En revanche, ressource complètement originale et manuellement développée par une équipe de lexicographes, ainsi qu'expliqué dans la section 1, il se distingue de ressources construites pour la désambiguïsation, plus ou moins automatiquement, en ré-utilisant et en enrichissant des ressources existantes (en particulier, des réseaux de type WordNet comme le font (Vossen *et al.*, 2011) ou (Tsatsaronis *et al.*, 2008)). L'annotation est riche de toute la richesse du RL-fr. Pour un mot-forme annoté du corpus RL-fr, le RL-fr fournit l'ensemble des copolysèmes entre lesquels désambiguïser ainsi que des informations linguistiques formalisées pour chaque copolysème (ex. caractéristiques grammaticales, étiquette sémantique, synonymes, contraires et de nombreux autres liens sémantiques ou de cooccurrence, encodés avec le système des fonctions lexicales, etc.).

Du point de vue de la désambiguïsation sémantique automatique, le corpus annoté RL-fr ressemble à un corpus d'évaluation (ou d'entraînement), puisque certains mots-formes du corpus RL-fr, relatifs à des vocables polysémiques donc ambigus dans une analyse automatique, sont *désambiguïsés* par une annotation avec une acception particulière¹⁴. Le corpus RL-fr se présente donc comme une ressource potentielle pour la *désambiguïsation sémantique automatique ciblée – targeted WSD*. Par contre, comme l'annotation du corpus RL-fr est relativement peu dense actuellement, le corpus n'est pas adapté pour la désambiguïsation sémantique globale – *all-words WSD*, (Navigli, 2009, p. 4).

Faisant l'hypothèse que plus l'annotation est dense plus le corpus annoté RL-fr est une ressource intéressante, nous encourageons la réutilisation des exemples dans RL-fr : un texte du corpus RL-fr peut être réutilisé et illustrer plusieurs unités lexicales, ce qui se traduit par l'annotation de plusieurs mots-formes de ce texte et donc, par une densification de l'annotation globale du corpus RL-fr. Pour faciliter ce processus, le corpus RL-fr est exporté chaque jour de la base de données RL-fr et importé dans TXM (Heiden, 2010). Les lexicographes disposent alors de toutes les fonctionnalités de la plateforme TXM pour leurs recherches dans le corpus-RL-fr.

Cependant, chaque texte du corpus est réutilisable en tant qu'exemple lexicographique *seulement jusqu'à un certain point* : un bon exemple E pour l'unité lexicale L₁ ne sera pas réutilisé pour illustrer l'unité lexicale L₂ s'il contient une occurrence ambiguë de L₂ (par exemple, par manque de contexte) ou une occurrence de L₂ sous des traits morpho-syntaxiques rares ou très particuliers, etc. Ainsi, la citation suivante :

À onze heures, avec son ami, il déjeune à la gargote d'une soupe, d'un plat de pommes de terre avec de la viande et de la salade. C'est un festin! (Frantext, FERAOUN Mouloud, *Le fils du pauvre*, 1950, p. 134)

illustre bien DÉJEUNER_V —le régime syntaxique *déjeuner de* est illustré, le second actant de l'unité lexicale est illustré avec plusieurs exemples de plats, etc. Mais le contexte est insuffisant pour savoir si *ami* signifie AMI_N 1.1 ('camarade') or AMI_N 1.2 ('petit ami') c'est pourquoi ce mot-forme ne peut pas être annoté.

Afin d'assurer une bonne qualité lexicographique du RL-fr, seules des occurrences *exemplaires* sont annotées dans le corpus, ce qui limite, *in fine*, la densité d'annotation du corpus-RL-fr. Ici, nos objectifs de lexicographie et les besoins de la désambiguïsation sémantique globale ne sont pas convergents.

Si des fins différentes des nôtres nécessitaient d'annoter plus densément le corpus RL-fr, il serait possible, dans notre architecture actuelle, d'associer à une unité lexicale L, un mot-forme dans un texte qui n'aurait pas le statut d'exemple lexicographique pour L – l'annotateur pourrait même préciser dans un commentaire sur l'association pourquoi le texte n'a pas valeur d'exemple pour L. Cependant, l'ambiguïté sémantique lexicale étant un phénomène naturel dans la langue, si tous les mots-formes du corpus-RL-fr devaient être annotés, il faudrait au moins prévoir la possibilité d'annoter un mot-forme avec une disjonction de sens et non pas forcément avec un seul sens, comme c'est le cas maintenant.

14. Dans le corpus RL-fr, des mots-formes relatifs à des vocables monosémiques, donc non ambigus, sont aussi annotés, ce qui n'est sans doute pas le cas dans un corpus annoté spécifiquement pour la désambiguïsation sémantique automatique.

Conclusions et perspectives

Dans cet article, nous nous sommes concentrés sur la collection d'exemples lexicographiques qui est une des ressources intégrées dans le système lexical RL-fr. Nous avons ainsi illustré les possibilités de changements de perspective – remarquablement faciles – dans le système lexical du RL-fr. Nous avons montré que cette collection d'exemple s'apparente à un corpus annoté potentiellement intéressant hors du projet RL-fr pour des recherches sur la désambiguïsation sémantique automatique. Nous avons analysé aussi la mesure dans laquelle nos intérêts dans RL-fr et ceux de cette communauté de recherche sont convergents pour l'annotation du corpus.

L'expérience de l'utilisation du corpus annoté pour entraîner ou évaluer un système de désambiguïsation reste encore à réaliser. Des questions pratiques se poseront que nous avons ignorées jusqu'ici : à quel type de système de désambiguïsation le corpus annoté convient-il ? le nombre d'entités annotées est-il suffisant pour entraîner un système statistique ? etc.

Et l'utilisation du réseau lexical RL-fr pour la désambiguïsation sémantique automatique se présente aussi comme une piste de recherche, d'autant plus intéressante que la désambiguïsation est une difficulté typique de l'analyse automatique, alors que la Lexicographie Explicative et Combinatoire se présente comme orientée vers la synthèse.

Remerciements

Merci à E. Jacquy, S. Ollinger, S. Pescarini, A. Polguère, D. Sikora et trois évaluateurs anonymes pour leurs commentaires constructifs sur une version précédente de cet article. Toutes les erreurs et insuffisances restent bien entendu de ma responsabilité.

Le projet RL-fr/RELIEF est financièrement soutenu par l'Agence de Mobilisation Économique de Lorraine (AMEL) et le Fonds Européen de Développement Régional (FEDER).

Références

- ATKINS S. & RUNDELL M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford : Oxford University Press.
- BAKER C., FILLMORE C. J. & LOWE J. (1998). The Berkeley Framenet Project. In *Proceedings of the COLING-ACL'98 Conference*, p. 86–90, Montreal, Canada.
- BAKER C. F., FILLMORE C. J. & CRONIN B. (2003). The Structure of the FrameNet Database. *International Journal of Lexicography*, **16**(3), 281–296.
- BARONI M., BERNARDINI S., FERRARESI A. & ZANCHETTA E. (2009). The WaCky Wide Web : A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation*, **43**(3), 209–226.
- BIRD S. & LIBERMAN M. (2001). A formal framework for linguistic annotation. *Speech Communication*, **33**(1-2), 23–60.
- DENDIEN J. & PIERREL J.-M. (2003). Le Trésor de la Langue Française informatisé : un exemple d'informatisation d'un dictionnaire de langue de référence. *Traitement Automatique des Langues (T.a.l.)*, **44**(2), 11–37.
- C. FELLBAUM, Ed. (1998). *WordNet : An Electronic Lexical Database*. Cambridge, MA : The MIT Press.
- GADER N., LUX-POGODALLA V. & POLGUÈRE A. (2012). Hand-crafting a lexical network with a knowledge based graph editor. In *Proceedings of CogALex-III Workshop of the 24th International Conference on Computational Linguistics (COLING2012)*, p. 109–125, Mumbai.
- HEIDEN S. (2010). The TXM platform : Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme. In *Proceedings of the 24th Pacific Asia Conference on Language Information and Computation*, Sendai, Japan : ENS-Lyon.
- KILGARRIFF A. & KOSEM I. (2012). Corpus tools for lexicographers. In S. GRANGER & M. PAQUOT, Eds., *Electronic lexicography*, p. 31–155. Oxford : Oxford University Press.
- KILGARRIFF A., RUNDELL M. & DHONNCHADHA E. (2006). Efficient corpus development for lexicography : building the New Corpus for Ireland. *Language Resources and Evaluation*, **40**(2), 127–152.

- LUX-POGODALLA V. & POLGUÈRE A. (2011). Construction of a French Lexical Network : Methodological Issues. In *Proceedings of WoLeR 2011 International Workshop on Linguistic Resources, ESSLLI (European Summer School in Logic, Language and Information, Ljubljana)*.
- MEL'ČUK I. (1996). Lexical Functions : A Tool for the Description of Lexical Relations in the Lexicon. In L. WANNER, Ed., *Lexical Functions in Lexicography and Natural Language Processing*, p. 37–102. Amsterdam/Philadelphia : Benjamins.
- MEL'ČUK I. (2006). Explanatory Combinatorial Dictionary. In G. SICA, Ed., *Open Problems in Linguistics and Lexicography*, p. 225–355. Monza : Polimetrica.
- MEL'ČUK I., CLAS A. & POLGUÈRE A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Paris/Louvain-la-Neuve : Duculot.
- MEL'ČUK I. & POLGUÈRE A. (2007). *Lexique actif du français. L'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*. Champs linguistiques. Brussels : De Boeck & Larcier.
- MEL'ČUK ET AL. I. (1984, 1988, 1992, 1999). *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques. Volumes I–IV*. Montréal : Les Presses de l'Université de Montréal.
- NAVIGLI R. (2009). Word sense disambiguation : a survey. *ACM Computing Surveys*, **41**(2), 1–69.
- POLGUÈRE A. (2009). Lexical systems : graph models of natural language lexicons. *Language Resources and Evaluation*, **43**(1), 41–55. Springer.
- POLGUÈRE A. (2012). Like a lexicographer weaving her lexical network. In *Proceedings of CogALex-III Workshop of the 24th International Conference on Computational Linguistics (COLING2012)*, p. 1–4, Mumbai.
- POLGUÈRE A. (2014). Principes de modélisation systématique des réseaux lexicaux. In *Actes de la 21ème conférence Traitement Automatique des Langues Naturelles*, Marseille.
- POLGUÈRE A. & SIKORA D. (2013). Modèle lexicographique de croissance du vocabulaire fondé sur un processus aléatoire, mais systématique. In C. R. C. MASSERON, C. GARCIA-DEBANC, Ed., *Enseigner le lexique. Pratiques sociales, objets à enseigner et pratiques d'enseignement*, collection de l'AiRDF 5. Laval, Québec : Presses de l'Université Laval.
- RAKHO M., ÉRIC LAPORTE & CONSTANT M. (2012). A new semantically annotated corpus with syntactic-semantic and cross-lingual senses. In N. C. C. CHAIR, K. CHOUKRI, T. DECLERCK, M. U. DOĞAN, B. MAEGAARD, J. MARIANI, A. MORENO, J. ODIJK & S. PIPERIDIS, Eds., *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey : European Language Resources Association (ELRA).
- SELVA T., VERLINDE S. & BINON J. (2003). Vers une deuxième génération de dictionnaires électroniques. *Traitement Automatique des Langues (T.A.L.)*, **44**(2), 177–197.
- SELVA T., VERLINDE S. & BINON J. (2009). Les bases de données au service d'un dictionnaire d'(auto-)apprentissage pour allophones. *Lexique*, **19**, 181–214.
- STEINLIN J., KAHANE S. & POLGUÈRE A. (2005). Compiling a “classical” explanatory combinatorial lexicographic description into a relational database. In *Proceedings of the Second International Conference on the Meaning Text Theory (MTT'2005)*, p. 477–485, Moscow.
- TSATSARONIS G., VARLAMIS I. & VAZIRGIANNIS M. (2008). Word Sense Disambiguation with Semantic Networks. In P. S. ET AL., Ed., *Text, Speech and Dialog - proceedings of the 11th international conference TSD*, volume 5246 of *Lectures Notes in Computer Science*, p. 219–226, Berlin-Heidelberg : Springer-Verlag.
- VOSSEN P., GÖRÖG A., LAAN F., VAN GOMPEL M., IZQUIERDO R. & VAN DEN BOSCH A. (2011). Dutchsemcor : Building a semantically annotated corpus for dutch. In *Proceedings of eLex 2011*, p. 286–296.